Joint Source-Channel Coding Revisited: Information-Spectrum Approach †

Te Sun HAN‡

February 2, 2008

†This paper is an extended refinement of a part of Chapter 3 in the book Han [11].

‡Te Sun Han was with Graduate School of Information Systems, University of Electro-Communications, Chofugaoka 1-5-1, Chofu, Tokyo 182-8585, Japan. He is now visiting Department of Computer Science, Faculty of Science and Engineering, Waseda University, Room 902, Bld. 201 (Shinjuku Lamdax Building), Ohkubo 2-4-12, Shinjuku-ku, Tokyo 169-0072, Japan. E-mail: han@is.uec.ac.jp, han@aoni.waseda.jp

Abstract: Given a general source $\mathbf{V} = \{V^n\}_{n=1}^{\infty}$ with *countably infinite* source alphabet and a general channel $\mathbf{W} = \{W^n\}_{n=1}^{\infty}$ with arbitrary *abstract* channel input/channel output alphabets, we study the joint source-channel coding problem from the information-spectrum point of view. First, we generalize Feinstein's lemma (direct part) and Verdú-Han's lemma (converse part) so as to be applicable to the general joint source-channel coding problem. Based on these lemmas, we establish a sufficient condition as well as a necessary condition for the source $\mathbf{V}$ to be reliably transmissible over the channel $\mathbf{W}$ with asymptotically vanishing probability of error. It is shown that our sufficient condition is equivalent to the sufficient condition derived by Vembu, Verdú and Steinberg [9], whereas our necessary condition is shown to be stronger than or equivalent to the necessary condition derived by them. It turns out, as a direct consequence, that "*separation principle*" in a relevantly generalized sense holds for a wide class of sources and channels, as was shown in a quite dfifferent manner by Vembu, Verdú and Steinberg [9]. It should also be remarked that a nice duality is found between our necessary and sufficient conditions, whereas we cannot fully enjoy such a duality between the necessary condition and the sufficient condition by Vembu, Verdú and Steinberg [9]. In addition, we demonstrate a sufficient condition as well as a necessary condition for the ε-transmissibility $(0 \le \varepsilon < 1)$. Finally, the separation theorem of the traditional standard form is shown to hold for the class of sources and channels that satisfy the semi-strong converse property.

Index terms: general source, general channel, joint source-channel coding, separation theorem, information-spectrum, transmissibility, generalized Feinstein's lemma, generalized Verdú-Han's lemma

1 Introduction

Given a source $\mathbf{V} = \{V^n\}_{n=1}^{\infty}$ and a channel $\mathbf{W} = \{W^n\}_{n=1}^{\infty}$, *joint source-channel coding* means that the encoder maps the output from the source directly to the channel input (*one step encoding*), where the probability of decoding error is required to vanish as block-length n tends to ∞. In usual situations, however, the joint source-channel coding can be decomposed into separate *source coding* and *channel coding* (*two step encoding*). This two step encoding does not cause any disadvantages from the standpoint of asymptotically vanishing error probabilities, provided that the so-called *Separation Theorem* holds.

Typically, the traditional separation theorem, which we call the separation theorem in the *narrow sense*, states that if the infimum $R_f(\mathbf{V})$ of all achievable fixed-length coding rates for the source $\mathbf{V}$ is smaller than the capacity $C(\mathbf{W})$ for the channel $\mathbf{W}$, then the source $\mathbf{V}$ is reliably transmissible by two step encoding over the channel $\mathbf{W}$; whereas if $R_f(\mathbf{V})$ is larger than $C(\mathbf{W})$ then the reliable transmission is impossible. While the former statement is always true for any general source $\mathbf{V}$ and any general channel $\mathbf{W}$, the latter statement is *not* always true. Then, a very natural question may be raised for what class of sources and channels and in what sense the separation theorem holds in general.

Shannon [1] has first shown that the separation theorem holds for the class of stationary memoryless sources and channels. Since then, this theorem has received extensive attention by a number of researchers who have attempted to prove versions that apply to more and more general classes of sources and channels. Among others, for example, Dobrushin [4], Pinsker [5], and Hu [6] have studied the separation theorem problem in the framework of information-stable sources and channels.

Recently, on the other hand, Vembu, Verdú and Steinberg [9] have put forth this problem in a much more general information-spectrum context with general source $\mathbf{V}$ and general channel $\mathbf{W}$. From the viewpoint of information spectra, they have generalized the notion of separation theorem and shown that, usually in many cases even with $R_f(\mathbf{V}) > C(\mathbf{W})$, it is possible to reliably transmit the output of the source $\mathbf{V}$ over the channel $\mathbf{W}$. Furthermore, in terms of information spectra, they have established a sufficient condition for the transmissibility as well as a necessary condition. It should be noticed here that, in this general joint source-channel coding situation, what indeed matters is not the validity problem of the traditional type of separation theorems but the derivation problem of necessary and/or

sufficient conditions for the transmissibility from the information-spectrum point of view.

However, while their sufficient condition looks simple and significantly tight, their necessary condition does not look quite close to tight.

The present paper was mainly motivated by the reasonable question why the forms of these two conditions look rather very different from one another. First, in Section 3, the basic tools to answer this question are established, i.e., two fundamental lemmas: a generalization of Feinstein's lemma [2] and a generalization of Verdú-Han's lemma [8], which provide with the very basis for the key results to be stated in the subsequent sections. These lemmas are of *dualistic* information-spectrum forms, which is in nice accordance with the general joint source-channel coding framework. In Section 4, given a general source $\mathbf{V}$ and a general channel $\mathbf{W}$, we establish, in terms of information-spectra, a sufficient condition (*Direct theorem*) for the transmissibility as well as a necessary condition (*Converse theorem*). The forms of these two conditions are very close from each other, and "fairly" coincides with one another, provided that we dare disregard some relevant asymptotically vanishing term.

Next, we equivalently rewrite these conditions in the forms useful to see relations to the separation theorem. As a consequence, it turns out that a separation-theorem-like equivalent of our sufficient condition just coincides with the sufficient condition given by Vembu, Verdú and Steinberg [9], whereas a separation-theorem-like equivalent of our necessary condition is shown to be strictly stronger than or equivalent to the necessary condition given by them. Here it is pleasing to observe that a nice duality is found between our necessary and sufficient conditions, whereas we cannot fully enjoy such a duality between the necessary condition and the sufficient condition by Vembu, Verdú and Steinberg [9].

On the other hand, in Section 5, we demonstrate a sufficient condition as well as a necessary condition for the ε-transmissibility, which is the generalization of the sufficient condition as well as the necessary condition as was shown in Section 4. Finally, in Section 6, we restrict the class of sources and channels to those that satisfy the strong converse property (or, more generally, the semi-strong converse property) to show that the separation theorem in the traditional sense holds for this class.

2 Basic Notation and Definitions

In this preliminary section, we prepare the basic notation and definitions which will be used in the subsequent sections.

2.1 General Sources

Let us first give here the formal defintion of the general source. A *general sources* is defined as an infinite sequence $\mathbf{V} = \{V^n = (V_1^{(n)}, \cdots, V_n^{(n)})\}_{n=1}^{\infty}$ of n-dimensional random variables V^n where each component random variable $V_i^{(n)}$ $(1 \le i \le n)$ takes values in a *countably infinite* set $\mathcal{V}$ that we call the *source alphabet*. It should be noted here that each component of V^n may change depending on block length n. This implies that the sequence $\mathbf{V}$ is quite general in the sense that it may not satisfy even the consistency condition as usual processes, where the consistency condition means that for any integers m, n such that $m < n$ it holds that $V_i^{(m)} \equiv V_i^{(n)}$ for all $i = 1, 2, \cdots, m$. The class of sources thus defined covers a very wide range of sources including all nonstationary and/or nonergodic sources (cf. Han and Verdú [7]).

2.2 General Channels

The formal definition of a general channel is as follows. Let $\mathcal{X}, \mathcal{Y}$ be arbitrary *abstract* (*not* necessarily countable) sets, which we call the *input alphabet* and the *output alphabet*, respectively. A *general channel* is defined as an infinite sequence $\mathbf{W} = \{W^n : \mathcal{X}^n \to \mathcal{Y}^n\}_{n=1}^{\infty}$ of n-dimensional probability transition matrices W^n, where $W^n(\mathbf{y}|\mathbf{x})$ $(\mathbf{x} \in \mathcal{X}^n, \mathbf{y} \in \mathcal{Y}^n)$ denotes the conditonal probability of $\mathbf{y}$ given $\mathbf{x}$.* The class of channels thus defined covers a very wide range of channels including all nonstationary and/or nonergodic channels with arbitrary memory structures (cf. Han and Verdú [7]).

Remark 2.1 A more reasonable definition of a general source is the following. Let $\{\mathcal{V}_n\}_{n=1}^{\infty}$ be any sequence of *arbitrary* source alphabets $\mathcal{V}_n$ (a countabley infinite or abstract set) and let V_n be any random variable taking values in $\mathcal{V}_n$ $(n = 1, 2, \cdots)$. Then, the sequence $\mathbf{V} = \{V_n\}_{n=1}^{\infty}$ of random

*In the case where the output alphabet $\mathcal{Y}$ is *abstract*, $W^n(\mathbf{y}|\mathbf{x})$ is understood to be the (conditional) probability measure element $W^n(d\mathbf{y}|\mathbf{x})$ that is measurable in $\mathbf{x}$.

variables V_n is called a *general source* (cf. Verdú and Han [10]). The above definition is a special case of this general source with $\mathcal{V}_n = \mathcal{V}^n$ $(n = 1, 2, \cdots)$.

On the other hand, a more reasonable definition of the general channel is the following. Let $\{W_n : \mathcal{X}_n \to \mathcal{Y}_n\}_{n=1}^{\infty}$ be any sequence of *arbitrary* probability transition matrices, where $\mathcal{X}_n, \mathcal{Y}_n$ are arbitrary abstract sets. Then, the sequence $\mathbf{W} = \{W_n\}_{n=1}^{\infty}$ of probability transition matrices W_n is called a *general channel* (cf. Han [11]). The above definition is a special case of this general channel with $\mathcal{X}_n = \mathcal{X}^n, \mathcal{Y}_n = \mathcal{Y}^n$ $(n = 1, 2, \cdots)$.

The results in this paper (Lemma 3.1, Lemma 3.2, Theorem 4.1, Theorem 4.2, Theorem 4.3, Theorem 4.4, Theorem 5.1, Theorem 5.2 and Theorems 6.1 $\sim$ 6.7) continue to be valid as well also in this more general setting with $\mathcal{V}^n, V^n, \mathbf{V}$ and $\mathcal{X}^n, \mathcal{Y}^n, W^n, \mathbf{W}$ replaced by $\mathcal{V}_n$, $V_n, \mathbf{V}$ and $\mathcal{X}_n, \mathcal{Y}_n, W_n, \mathbf{W}$, respectively.

In the sequel we use the convention that $P_Z(\cdot)$ denotes the probability distribution of a random variable Z, whereas $P_{Z|U}(\cdot|\cdot)$ denotes the conditional probability distribution of a random variable Z given a random variable U. $\square$

2.3 Joint Source-Channel Coding

Let $\mathbf{V} = \{V^n = (V_1^{(n)}, \cdots, V_n^{(n)})\}_{n=1}^{\infty}$ be any general source, and let $\mathbf{W} = \{W^n(\cdot|\cdot) : \mathcal{X}^n \to \mathcal{Y}^n\}_{n=1}^{\infty}$ be any general channel. We consider an *encoder* $\varphi_n : \mathcal{V}^n \to \mathcal{X}^n$ and a *decoder* $\psi_n : \mathcal{Y}^n \to \mathcal{V}^n$, and put $X^n = \varphi_n(V^n)$. Then, denoting by Y^n the output from the channel W^n due to the input X^n, we have the obvious relation:

$$V^n \to X^n \to Y^n \quad \text{(a Markov chain)}. \tag{2.1}$$

The *error probability* ε_n with code (φ_n, ψ_n) is defined by

$$\begin{aligned} \varepsilon_n &\equiv \Pr\{V^n \neq \psi_n(Y^n)\} \\ &= \sum_{\mathbf{v} \in \mathcal{V}^n} P_{V^n}(\mathbf{v}) W^n(\mathcal{D}^c(\mathbf{v})|\varphi_n(\mathbf{v})), \end{aligned} \tag{2.2}$$

where $\mathcal{D}(\mathbf{v}) \equiv \{\mathbf{y} \in \mathcal{Y}^n | \psi_n(\mathbf{y}) = \mathbf{v}\}$ $(\forall \mathbf{v} \in \mathcal{V}^n)$ ($\mathcal{D}(\mathbf{v})$ is called the *decoding set* for $\mathbf{v}$) and "c" denotes the complement of a set. A pair (φ_n, ψ_n) with error probability ε_n is simply called a joint source-channel code (n, ε_n).

We now define the *transmissibility* in terms of joint source-channel codes (n, ε_n) as

Definition 2.1

Source $\mathbf{V}$ is transmissible over channel $\mathbf{W}$ $\overset{\text{def}}{\Longleftrightarrow}$ There exists an (n, ε_n) code such that $\lim_{n\to\infty} \varepsilon_n = 0$.

With this definition of transmissibility, in the following sections we shall establish a sufficient condition as well as a necessary condition for the transmissibility when we are given a geneal source $\mathbf{V}$ and a general channel $\mathbf{W}$. These two conditions are *very close* to each other and could actually be seen as giving "*almost the same condition,*" provided that we dare disregard an asymptotically negligible term $\gamma_n \to 0$ appearing in those conditions (cf. Section 4).

Remark 2.2 The quantity ε_n defined by (2.2) is more specifically called the *average* error probability, because it is averaged with respect to $P_{V^n}(\mathbf{v})$ over all source outputs $\mathbf{v} \in \mathcal{V}^n$. On the other hand, we may define another kind of error probability by

$$\varepsilon_n \equiv \sup_{\mathbf{v}:P_{V^n}(\mathbf{v})>0} W^n(\mathcal{D}^c(\mathbf{v})|\varphi_n(\mathbf{v})), \tag{2.3}$$

which we call the *maximum* error probability. It is evident that the transmissibility in the maximum sense implies the transmissibility in the average sense. However, the inverse is not necessarily true. To see this, it suffices to consider the following simple example. Let the source, channel input, channel output alphabets be $\mathcal{V}_n = \{0,1,2\}$, $\mathcal{X}_n = \{1,2\}$, $\mathcal{Y}_n = \{1,2\}$, respectively; and the (deterministic) channel $W_n : \mathcal{X}_n \to \mathcal{Y}_n$ be defined by $W_n(j|i) = 1$ for $i = j, W_n(1|0) = 1$. Moreover, let the source V_n have probability distribution $P_{V_n}(0) = \alpha_n,\ P_{V_n}(1) = P_{V_n}(2) = \frac{1-\alpha_n}{2}$ ($\alpha_n \to 0$ as $n \to \infty$). One of the best choices of possible pairs of encoder-decoder ($\varphi_n : \mathcal{V}_n \to \mathcal{X}_n,\ \psi_n : \mathcal{Y}_n \to \mathcal{V}_n$), either in the average sense or in the maximum sense, is such that $\varphi_n(i) = i$ for $i = 1,2; \varphi_n(0) = 1;\ \psi_n(i) = i$ for $i = 1,2$. Then, the average error probability is $\varepsilon_n^{\mathrm{a}} = \alpha_n \to 0$, while the maximum error probability is $\varepsilon_n^{\mathrm{m}} = 1$. Thus, in this case, the source V_n is transmissible in the average sense over the channel W_n, while it is *not* transmissible in the maximum sense.

Hereafter, the probability ε_n is understood to denote the "average" error probability, unless otherwise stated. $\square$

3 Fundamental Lemmas

In this section, we prepare two fundamental lemmas that are needed in the next section in order to establish the main theorems (*Direct part* and *Converse part*).

Lemma 3.1 (Generalization of Feinstein's lemma) Given a general source $\mathbf{V} = \{V^n\}_{n=1}^{\infty}$ and a general channel $\mathbf{W} = \{W^n\}_{n=1}^{\infty}$, let X^n be any input random variable taking values in $\mathcal{X}^n$ and Y^n be the channel output via W^n due to the channel input X^n, where $V^n \to X^n \to Y^n$. Then, for every $n = 1, 2, \cdots$, there exists an (n, ε_n) code such that

$$\varepsilon_n \le \Pr\left\{\frac{1}{n}\log\frac{W^n(Y^n|X^n)}{P_{Y^n}(Y^n)} \le \frac{1}{n}\log\frac{1}{P_{V^n}(V^n)} + \gamma\right\} + e^{-n\gamma}, \tag{3.1}$$

where[†] $\gamma > 0$ is an arbitrary positive number.

Remark 3.1 In a special case where the source $\mathbf{V} = \{V^n\}_{n=1}^{\infty}$ is uniformly distributed on the massage set $\mathcal{M}_n = \{1, 2, \cdots, M_n\}$, it follows that

$$\frac{1}{n}\log\frac{1}{P_{V^n}(V^n)} = \frac{1}{n}\log M_n,$$

which implies that the entropy spectrum[‡] of the source $\mathbf{V} = \{V^n\}_{n=1}^{\infty}$ is exactly one point spectrum concentrated on $\frac{1}{n}\log M_n$. Therefore, in this special case, Lemma 3.1 reduecs to Feinstein's lemma [2]. □

Proof of Lemma3.1:

For each $\mathbf{v} \in \mathcal{V}^n$, generate $\mathbf{x}(\mathbf{v}) \in \mathcal{X}^n$ at random according to the conditional distribution $P_{X^n|V^n}(\cdot|\mathbf{v})$ and let $\mathbf{x}(\mathbf{v})$ be the codeword for $\mathbf{v}$. In other words, we define the encoder $\varphi_n : \mathcal{V}^n \to \mathcal{X}^n$ as $\varphi_n(\mathbf{v}) = \mathbf{x}(\mathbf{v})$, where

[†]In the case where the input and output alphabets $\mathcal{X}, \mathcal{Y}$ are *abstract* (*not* necessarily countable), $\frac{W^n(Y^n|X^n)}{P_{Y^n}(Y^n)}$ in (3.1) is understood to be $g(Y^n|X^n)$, where $g(\mathbf{y}|\mathbf{x}) \equiv \frac{W^n(d\mathbf{y}|\mathbf{x})}{P_{Y^n}(d\mathbf{y})}$ $= \frac{W^n(d\mathbf{y}|\mathbf{x})P_{X^n}(d\mathbf{x})}{P_{Y^n}(d\mathbf{y})P_{X^n}(d\mathbf{x})} = \frac{P_{X^nY^n}(d\mathbf{x},d\mathbf{y})}{P_{X^n}(d\mathbf{x})P_{Y^n}(d\mathbf{y})}$ is the Radon-Nikodym derivative that is measurable in $(\mathbf{x}, \mathbf{y})$.

[‡]The probablity distribution of $\frac{1}{n}\log\frac{1}{P_{V^n}(V^n)}$ is called the *entropy spectrum* of the source $\mathbf{V} = \{V^n\}_{n=1}^{\infty}$, whereas the probability distribution of $\frac{1}{n}\log\frac{W^n(Y^n|X^n)}{P_{Y^n}(Y^n)}$ is called the *mutual information spectrum* of the channel $\mathbf{W} = \{W^n\}_{n=1}^{\infty}$ given the input $\mathbf{X} = \{X^n\}_{n=1}^{\infty}$ (cf. Han and Verdú [7]).

$\{\mathbf{x}(\mathbf{v}) \mid \forall \mathbf{v} \in \mathcal{V}^n\}$ are all independently generated. We define the decoder $\psi_n : \mathcal{Y}^n \to \mathcal{V}^n$ as follows: Set

$$S_n = \left\{ (\mathbf{v}, \mathbf{x}, \mathbf{y}) \in \mathcal{Z}^n \left| \frac{1}{n} \log \frac{W^n(\mathbf{y}|\mathbf{x})}{P_{Y^n}(\mathbf{y})} > \frac{1}{n} \log \frac{1}{P_{V^n}(\mathbf{v})} + \gamma \right. \right\}, \tag{3.2}$$

$$S_n(\mathbf{v}) = \{(\mathbf{x}, \mathbf{y}) \in \mathcal{X}^n \times \mathcal{Y}^n \,|(\mathbf{v}, \mathbf{x}, \mathbf{y}) \in S_n\}, \tag{3.3}$$

where for simplicity we have put $\mathcal{Z}^n \equiv \mathcal{V}^n \times \mathcal{X}^n \times \mathcal{Y}^n$. Suppose that the decoder ψ_n received a channel output $\mathbf{y} \in \mathcal{Y}^n$. If there exists one and only one $\mathbf{v} \in \mathcal{V}^n$ such that $(\mathbf{x}(\mathbf{v}), \mathbf{y}) \in S_n(\mathbf{v})$, define the decoder as $\psi_n(\mathbf{y}) = \mathbf{v}$; otherwise, let the output of the decoder $\psi_n(\mathbf{y}) \in \mathcal{V}^n$ be arbitrary. Then, the probability $\overline{\varepsilon}_n$ of error for this pair (φ_n, ψ_n) (averaged over all the realizatioins of the random code) is given by

$$\overline{\varepsilon}_n = \sum_{\mathbf{v} \in \mathcal{V}^n} P_{V^n}(\mathbf{v}) \overline{\varepsilon}_n(\mathbf{v}), \tag{3.4}$$

where $\overline{\varepsilon}_n(\mathbf{v})$ is the probability of error (averaged over all the realizatioins of the random code) when $\mathbf{v} \in \mathcal{V}^n$ is the source output. We can evaluate $\overline{\varepsilon}_n(\mathbf{v})$ as

$$\begin{aligned} \overline{\varepsilon}_n(\mathbf{v}) &\le \Pr\{(\mathbf{x}(\mathbf{v}), Y^n) \notin S_n(\mathbf{v})\} \\ &\quad + \Pr\left\{ \bigcup_{\mathbf{v}' : \mathbf{v}' \neq \mathbf{v}} \{(\mathbf{x}(\mathbf{v}'), Y^n) \in S_n(\mathbf{v}')\} \right\} \\ &\le \Pr\{(\mathbf{x}(\mathbf{v}), Y^n) \notin S_n(\mathbf{v})\} \\ &\quad + \sum_{\mathbf{v}' : \mathbf{v}' \neq \mathbf{v}} \Pr\{(\mathbf{x}(\mathbf{v}'), Y^n) \in S_n(\mathbf{v}')\}, \end{aligned} \tag{3.5}$$

where Y^n is the channel output via W^n due to the channel input $\mathbf{x}(\mathbf{v})$. The first term on the right-hand side of (3.5) is written as

$$\begin{aligned} A_n(\mathbf{v}) &\equiv \Pr\{(\mathbf{x}(\mathbf{v}), Y^n) \notin S_n(\mathbf{v})\} \\ &= \sum_{(\mathbf{x}, \mathbf{y}) \notin S_n(\mathbf{v})} P_{X^n Y^n | V^n}(\mathbf{x}, \mathbf{y} | \mathbf{v}). \end{aligned}$$

Hence,

$$\begin{aligned} \sum_{\mathbf{v} \in \mathcal{V}^n} P_{V^n}(\mathbf{v}) A_n(\mathbf{v}) &= \sum_{\mathbf{v} \in \mathcal{V}^n} P_{V^n}(\mathbf{v}) \sum_{(\mathbf{x}, \mathbf{y}) \notin S_n(\mathbf{v})} P_{X^n Y^n | V^n}(\mathbf{x}, \mathbf{y} | \mathbf{v}) \\ &= \sum_{(\mathbf{v}, \mathbf{x}, \mathbf{y}) \notin S_n} P_{V^n X^n Y^n}(\mathbf{v}, \mathbf{x}, \mathbf{y}) \\ &= \Pr\{V^n X^n Y^n \notin S_n\}. \end{aligned} \tag{3.6}$$

On the other hand, noting that $\mathbf{x}(\mathbf{v}'), \mathbf{x}(\mathbf{v})$ ($\mathbf{v}' \neq \mathbf{v}$) are independent and hence $\mathbf{x}(\mathbf{v}')$, Y^n are also independent, the second term on the right-hand side of (3.5) is evaluated as

$$\begin{aligned} B_n(\mathbf{v}) &\equiv \sum_{\mathbf{v}':\mathbf{v}'\neq\mathbf{v}} \Pr\left\{(\mathbf{x}(\mathbf{v}'), Y^n) \in S_n(\mathbf{v}')\right\} \\ &= \sum_{\mathbf{v}':\mathbf{v}'\neq\mathbf{v}} \sum_{(\mathbf{x},\mathbf{y})\in S_n(\mathbf{v}')} P_{Y^n|V^n}(\mathbf{y}|\mathbf{v}) P_{X^n|V^n}(\mathbf{x}|\mathbf{v}') \\ &\leq \sum_{\mathbf{v}'\in\mathcal{V}^n} \sum_{(\mathbf{x},\mathbf{y})\in S_n(\mathbf{v}')} P_{Y^n|V^n}(\mathbf{y}|\mathbf{v}) P_{X^n|V^n}(\mathbf{x}|\mathbf{v}'). \end{aligned}$$

Hence,

$$\begin{aligned} &\sum_{\mathbf{v}\in\mathcal{V}^n} P_{V^n}(\mathbf{v}) B_n(\mathbf{v}) \\ &\leq \sum_{\mathbf{v}\in\mathcal{V}^n} \sum_{\mathbf{v}'\in\mathcal{V}^n} \sum_{(\mathbf{x},\mathbf{y})\in S_n(\mathbf{v}')} P_{V^n}(\mathbf{v}) P_{Y^n|V^n}(\mathbf{y}|\mathbf{v}) P_{X^n|V^n}(\mathbf{x}|\mathbf{v}') \\ &= \sum_{\mathbf{v}'\in\mathcal{V}^n} \sum_{(\mathbf{x},\mathbf{y})\in S_n(\mathbf{v}')} P_{Y^n}(\mathbf{y}) P_{X^n|V^n}(\mathbf{x}|\mathbf{v}'). \end{aligned} \tag{3.7}$$

On the other hand, in view of (3.2), (3.3), $(\mathbf{x}, \mathbf{y}) \in S_n(\mathbf{v}')$ implies

$$P_{Y^n}(\mathbf{y}) \leq P_{V^n}(\mathbf{v}') W^n(\mathbf{y}|\mathbf{x}) e^{-n\gamma}.$$

Therefore, (3.7) is further transformed to

$$\begin{aligned} &\sum_{\mathbf{v}\in\mathcal{V}^n} P_{V^n}(\mathbf{v}) B_n(\mathbf{v}) \\ &\leq e^{-n\gamma} \sum_{\mathbf{v}'\in\mathcal{V}^n} \sum_{(\mathbf{x},\mathbf{y})\in S_n(\mathbf{v}')} P_{V^n}(\mathbf{v}') P_{X^n|V^n}(\mathbf{x}|\mathbf{v}') W^n(\mathbf{y}|\mathbf{x}) \\ &\leq e^{-n\gamma} \sum_{(\mathbf{v}',\mathbf{x},\mathbf{y})\in\mathcal{Z}^n} P_{V^n}(\mathbf{v}') P_{X^n|V^n}(\mathbf{x}|\mathbf{v}') W^n(\mathbf{y}|\mathbf{x}) \\ &= e^{-n\gamma}. \end{aligned} \tag{3.8}$$

Then, from (3.4), (3.6) and (3.8) it follows that

$$\begin{aligned} \overline{\varepsilon}_n &= \sum_{\mathbf{v}\in\mathcal{V}^n} P_{V^n}(\mathbf{v}) \overline{\varepsilon}_n(\mathbf{v}) \\ &\leq \sum_{\mathbf{v}\in\mathcal{V}^n} P_{V^n}(\mathbf{v}) A_n(\mathbf{v}) + \sum_{\mathbf{v}\in\mathcal{V}^n} P_{V^n}(\mathbf{v}) B_n(\mathbf{v}) \\ &\leq \Pr\left\{V^n X^n Y^n \notin S_n\right\} + e^{-n\gamma}. \end{aligned}$$

Thus, there must exist a deterministic (n, ε_n) code such that

$$\varepsilon_n \le \Pr\left\{V^n X^n Y^n \notin S_n\right\} + e^{-n\gamma},$$

thereby proving Lemma 3.1. □

Lemma 3.2 (Generalization of Verdú-Han's lemma) Let $\mathbf{V} = \{V^n\}_{n=1}^{\infty}$ and $\mathbf{W} = \{W^n\}_{n=1}^{\infty}$ be a general source and a general channel, respectively, and let $\varphi_n : \mathcal{V}^n \to \mathcal{X}^n$ be the encoder of an (n, ε_n) code for (V^n, W^n). Put $X^n = \varphi_n(V^n)$ and let Y^n be the channel output via W^n due to the channel input X^n, where $V^n \to X^n \to Y^n$. Then, for every $n = 1, 2, \cdots$, it holds that

$$\varepsilon_n \ge \Pr\left\{\frac{1}{n}\log\frac{W^n(Y^n|X^n)}{P_{Y^n}(Y^n)} \le \frac{1}{n}\log\frac{1}{P_{V^n}(V^n)} - \gamma\right\} - e^{-n\gamma}, \qquad (3.9)$$

where $\gamma > 0$ is an arbitrary positive number.

Remark 3.2 In a special case where the source $\mathbf{V} = \{V^n\}_{n=1}^{\infty}$ is uniformly distributed on the massage set $\mathcal{M}_n = \{1, 2, \cdots, M_n\}$, it follows that

$$\frac{1}{n}\log\frac{1}{P_{V^n}(V^n)} = \frac{1}{n}\log M_n,$$

which implies that the entropy spectrum of the source $\mathbf{V} = \{V^n\}_{n=1}^{\infty}$ is exactly one point spectrum concentrated on $\frac{1}{n}\log M_n$. Therefore, in this special case, Lemma 3.2 reduecs to Verdú-Han's lemma [8]. □

Proof of Lemma3.2

Define

$$L_n = \left\{(\mathbf{v}, \mathbf{x}, \mathbf{y}) \in \mathcal{Z}^n \left| \frac{1}{n}\log\frac{W^n(\mathbf{y}|\mathbf{x})}{P_{Y^n}(\mathbf{y})} \le \frac{1}{n}\log\frac{1}{P_{V^n}(\mathbf{v})} - \gamma \right.\right\}, \qquad (3.10)$$

and, for each $\mathbf{v} \in \mathcal{V}^n$ set

$$\mathcal{D}(\mathbf{v}) = \left\{\mathbf{y} \in \mathcal{Y}^n | \psi_n(\mathbf{y}) = \mathbf{v}\right\},$$

that is, $\mathcal{D}(\mathbf{v})$ is the decoding set for $\mathbf{v}$. Moreover, for each $(\mathbf{v}, \mathbf{x}) \in \mathcal{V}^n \times \mathcal{X}^n$, set

$$\mathcal{B}(\mathbf{v}, \mathbf{x}) = \left\{\mathbf{y} \in \mathcal{Y}^n | (\mathbf{v}, \mathbf{x}, \mathbf{y}) \in L_n\right\}. \qquad (3.11)$$

Then, noting the Markov chain property (2.1), we have

$$
\begin{aligned}
&\Pr\{V^nX^nY^n \in L_n\}\\
&= \sum_{(\mathbf{v},\mathbf{x},\mathbf{y})\in L_n} P_{V^nX^nY^n}(\mathbf{v},\mathbf{x},\mathbf{y})\\
&= \sum_{(\mathbf{v},\mathbf{x})\in\mathcal{V}^n\times\mathcal{X}^n} P_{V^nX^n}(\mathbf{v},\mathbf{x})W^n(\mathcal{B}(\mathbf{v},\mathbf{x})|\mathbf{x})\\
&= \sum_{(\mathbf{v},\mathbf{x})\in\mathcal{V}^n\times\mathcal{X}^n} P_{V^nX^n}(\mathbf{v},\mathbf{x})W^n(\mathcal{B}(\mathbf{v},\mathbf{x})\cap\mathcal{D}^c(\mathbf{v})|\mathbf{x})\\
&\quad + \sum_{(\mathbf{v},\mathbf{x})\in\mathcal{V}^n\times\mathcal{X}^n} P_{V^nX^n}(\mathbf{v},\mathbf{x})W^n(\mathcal{B}(\mathbf{v},\mathbf{x})\cap\mathcal{D}(\mathbf{v})|\mathbf{x})\\
&\le \sum_{(\mathbf{v},\mathbf{x})\in\mathcal{V}^n\times\mathcal{X}^n} P_{V^nX^n}(\mathbf{v},\mathbf{x})W^n(\mathcal{D}^c(\mathbf{v})|\mathbf{x})\\
&\quad + \sum_{(\mathbf{v},\mathbf{x})\in\mathcal{V}^n\times\mathcal{X}^n} P_{V^nX^n}(\mathbf{v},\mathbf{x})W^n(\mathcal{B}(\mathbf{v},\mathbf{x})\cap\mathcal{D}(\mathbf{v})|\mathbf{x})\\
&= \varepsilon_n + \sum_{(\mathbf{v},\mathbf{x})\in\mathcal{V}^n\times\mathcal{X}^n} P_{V^nX^n}(\mathbf{v},\mathbf{x})W^n(\mathcal{B}(\mathbf{v},\mathbf{x})\cap\mathcal{D}(\mathbf{v})|\mathbf{x})\\
&= \varepsilon_n + \sum_{(\mathbf{v},\mathbf{x})\in\mathcal{V}^n\times\mathcal{X}^n} P_{V^nX^n}(\mathbf{v},\mathbf{x}) \sum_{\mathbf{y}\in\mathcal{B}(\mathbf{v},\mathbf{x})\cap\mathcal{D}(\mathbf{v})} W^n(\mathbf{y}|\mathbf{x}), \qquad (3.12)
\end{aligned}
$$

where we have used the relation:

$$
\varepsilon_n = \sum_{(\mathbf{v},\mathbf{x})\in\mathcal{V}^n\times\mathcal{X}^n} P_{V^nX^n}(\mathbf{v},\mathbf{x})W^n(\mathcal{D}^c(\mathbf{v})|\mathbf{x}).
$$

Now, it follows from (3.10) and (3.11) that $\mathbf{y}\in\mathcal{B}(\mathbf{v},\mathbf{x})$ implies

$$
W^n(\mathbf{y}|\mathbf{x}) \le \frac{e^{-n\gamma}P_{Y^n}(\mathbf{y})}{P_{V^n}(\mathbf{v})},
$$

which is substituted into the right-hand side of (3.12) to yield

$$
\begin{aligned}
&\Pr\{V^nX^nY^n \in L_n\}\\
&\le \varepsilon_n + e^{-n\gamma} \sum_{(\mathbf{v},\mathbf{x})\in\mathcal{V}^n\times\mathcal{X}^n} P_{X^n|V^n}(\mathbf{x}|\mathbf{v}) \sum_{\mathbf{y}\in\mathcal{B}(\mathbf{v},\mathbf{x})\cap\mathcal{D}(\mathbf{v})} P_{Y^n}(\mathbf{y})\\
&\le \varepsilon_n + e^{-n\gamma} \sum_{(\mathbf{v},\mathbf{x})\in\mathcal{V}^n\times\mathcal{X}^n} P_{X^n|V^n}(\mathbf{x}|\mathbf{v})P_{Y^n}(\mathcal{D}(\mathbf{v}))\\
&= \varepsilon_n + e^{-n\gamma} \sum_{\mathbf{v}\in\mathcal{V}^n} P_{Y^n}(\mathcal{D}(\mathbf{v}))\\
&= \varepsilon_n + e^{-n\gamma},
\end{aligned}
$$

thereby proving the claim of the lemma. $\square$

4 Theorems on Transmissibility

In this section we give both of a sufficient condition and a necessary condition for the transmissibility with a given general souce $\mathbf{V} = \{V^n\}_{n=1}^{\infty}$ and a given general channel $\mathbf{W} = \{W^n\}_{n=1}^{\infty}$.

First, Lemma 3.1 immediately leads us to the following direct theorem:

Theorem 4.1 (Direct theorem) Let $\mathbf{V} = \{V^n\}_{n=1}^{\infty}$, $\mathbf{W} = \{W^n\}_{n=1}^{\infty}$ be a general source and a general channel, respectively. If there exist *some* channel input $\mathbf{X} = \{X^n\}_{n=1}^{\infty}$ and *some* sequence $\{\gamma_n\}_{n=1}^{\infty}$ satisfying

$$\gamma_n > 0,\ \gamma_n \to 0 \text{ and } n\gamma_n \to \infty \quad (n \to \infty) \tag{4.1}$$

for which it holds that

$$\lim_{n\to\infty} \Pr\left\{ \frac{1}{n} \log \frac{W^n(Y^n|X^n)}{P_{Y^n}(Y^n)} \le \frac{1}{n} \log \frac{1}{P_{V^n}(V^n)} + \gamma_n \right\} = 0, \tag{4.2}$$

then the source $\mathbf{V} = \{V^n\}_{n=1}^{\infty}$ is transmissible over the channel $\mathbf{W} = \{W^n\}_{n=1}^{\infty}$, where Y^n is the channel output via W^n due to the channel input X^n and $V^n \to X^n \to Y^n$.

Proof:

Since in Lemma 3.1 we can choose the constant $\gamma > 0$ so as to depend on n, let us take, instead of γ, an arbitrary $\{\gamma_n\}_{n=1}^{\infty}$ satisfying condition (4.1). Then, the second term on the right-hand side of (3.1) vanishes as n tends to ∞, and hence it follows from (4.2) that the right-hand side of (3.1) vanishes as n tends to ∞. Therefore, the (n, ε_n) code as specified in Lemma 3.1 satisfies $\lim\limits_{n\to\infty} \varepsilon_n = 0$. $\square$

Next, Lemma 3.2 immediately leads us to the following converse theorem:

Theorem 4.2 (Converse theorem) Suppose that a general source $\mathbf{V} = \{V^n\}_{n=1}^{\infty}$ is transmissible over a general channel $\mathbf{W} = \{W^n\}_{n=1}^{\infty}$. Let the channel input be $\mathbf{X} = \{X^n \equiv \varphi_n(V^n)\}_{n=1}^{\infty}$ where $\varphi_n : \mathcal{V}^n \to \mathcal{X}^n$ is the channel encoder. Then, for *any* sequence $\{\gamma_n\}_{n=1}^{\infty}$ satisfying condition (4.1), it holds that

$$\lim_{n\to\infty} \Pr\left\{ \frac{1}{n} \log \frac{W^n(Y^n|X^n)}{P_{Y^n}(Y^n)} \le \frac{1}{n} \log \frac{1}{P_{V^n}(V^n)} - \gamma_n \right\} = 0, \tag{4.3}$$

where Y^n is the channel output via W^n due to the channel input X^n and $V^n \to X^n \to Y^n$.

Proof:

If $\mathbf{V}$ is transmissible over $\mathbf{W}$, then, by Definition 2.1 there exists an (n, ε_n) code such that $\lim_{n\to\infty} \varepsilon_n = 0$. Hence, the claim of the theorem immediately follows from (3.9) in Lemma 3.2 with γ_n instead of γ. $\square$

Remark 4.1 Comparing (4.3) in Theorem 4.2 with (4.2) in Theorem 4.1, we observe that the only difference is that the sign of γ_n is changed from $+$ to $-$. Since γ_n vanishes as n tends to ∞, this difference is asymptotically negligible. $\square$

Now, let us think of the implication of conditions (4.2) and (4.3). First, let us think of (4.2). Putting

$$A_n = \frac{1}{n}\log\frac{W^n(Y^n|X^n)}{P_{Y^n}(Y^n)}, \quad B_n = \frac{1}{n}\log\frac{1}{P_{V^n}(V^n)}$$

for simplicity, (4.2) is written as

$$\alpha_n \equiv \Pr\left\{A_n \le B_n + \gamma_n\right\} \to 0 \quad (n \to \infty), \tag{4.4}$$

which can be transformed to

$$\begin{aligned}
&\Pr\left\{A_n \le B_n + \gamma_n\right\} \\
&\quad = \sum_u \Pr\left\{B_n = u\right\} \Pr\left\{A_n \le B_n + \gamma_n | B_n = u\right\} \\
&\quad = \sum_u \Pr\left\{B_n = u\right\} \Pr\left\{A_n \le u + \gamma_n | B_n = u\right\}.
\end{aligned}$$

Set

$$T_n = \left\{u \mid \Pr\left\{A_n \le u + \gamma_n | B_n = u\right\} \le \sqrt{\alpha_n}\right\}, \tag{4.5}$$

then by virtue of (4.4) and Markov inequality, we have

$$\Pr\left\{B_n \in T_n\right\} \ge 1 - \sqrt{\alpha_n}. \tag{4.6}$$

Let us now define the upper cumulative probabilities for A_n, B_n by

$$P_n(t) = \Pr\left\{A_n \ge t\right\}, \quad Q_n(t) = \Pr\left\{B_n \ge t\right\},$$

then it follows that

$$
\begin{aligned}
P_n itj &= \sum_{u} \Pr\{B_n = u\} \Pr\{A_n \geq t | B_n = u\} \\
&\geq \sum_{\substack{u \in T_n: \\ u \geq t-\gamma_n}} \Pr\{B_n = u\} \Pr\{A_n \geq t | B_n = u\} \\
&\geq \sum_{\substack{u \in T_n: \\ u \geq t-\gamma_n}} \Pr\{B_n = u\} \Pr\{A_n \geq u + \gamma_n | B_n = u\}. \qquad (4.7)
\end{aligned}
$$

On the other hand, by means of (4.5), $u \in T_n$ implies that

$$\Pr\{A_n \geq u + \gamma_n | B_n = u\} \geq 1 - \sqrt{\alpha_n}.$$

Theore, by (4.6), (4.7) it is concluded that

$$
\begin{aligned}
P_n(t) &\geq (1 - \sqrt{\alpha_n}) \sum_{\substack{u \in T_n: \\ u \geq t-\gamma_n}} \Pr\{B_n = u\} \\
&\geq (1 - \sqrt{\alpha_n})(Q_n(t - \gamma_n) - \Pr\{B_n \notin T_n\}) \\
&\geq (1 - \sqrt{\alpha_n})(Q_n(t - \gamma_n) - \sqrt{\alpha_n}) \\
&\geq Q_n(t - \gamma_n) - 2\sqrt{\alpha_n}.
\end{aligned}
$$

That is,

$$P_n(t) \geq Q_n(t - \gamma_n) - 2\sqrt{\alpha_n}.$$

This means that, for all t, the upper cumulative probability $P_n(t)$ of A_n is larger than or equal to the upper cumulative probability $Q_n(t - \gamma_n)$ of B_n, except for the asymptotically vanishing difference $2\sqrt{\alpha_n}$. This in turn implies that, as a whole, the mutual information spectrum of the channel is shifted to the right in comparison with the entropy spectrum of the source. With $-\gamma_n$ instead of γ_n, the same implication follows also from (4.3). It is such an allocation relation between the mutual information spectrum and the entropy spectrum that enables us to make an transmissible joint source-channel coding.

However, it is not easy in general to check whether conditions (4.2), (4.3) in these forms are satisfied or not. Therefore, we consider to equivalently rewrite conditions (4.2), (4.3) into alternative information-spectrum forms hopefully easier to depict an intuitive picture. This can actually be done by re-choosing the input and output variables X^n, Y^n as below. These forms

are useful in order to see the relation of conditions (4.2), (4.3) with the so-called *separation theorem*.

First, we show another information-spectrum form equivalent to the sufficient condition (4.2) in Theorem 4.1.

Theorem 4.3 (Equivalence of sufficient conditions) The following two conditions are equivalent:

1) For *some* channel input $\mathbf{X} = \{X^n\}_{n=1}^{\infty}$ and *some* sequence $\{\gamma_n\}_{n=1}^{\infty}$ satisfying condition (4.1), it holds that

$$\lim_{n\to\infty} \Pr\left\{\frac{1}{n}\log\frac{W^n(Y^n|X^n)}{P_{Y^n}(Y^n)} \le \frac{1}{n}\log\frac{1}{P_{V^n}(V^n)} + \gamma_n\right\} = 0, \tag{4.8}$$

where Y^n is the channel output via W^n due to the channel input X^n and $V^n \to X^n \to Y^n$.

2) (**Strict domination:** Vembu, Verdú and Steinberg [9]) For *some* channel input $\mathbf{X} = \{X^n\}_{n=1}^{\infty}$, *some* sequence $\{c_n\}_{n=1}^{\infty}$ and *some* sequence $\{\gamma_n\}_{n=1}^{\infty}$ satisfying condition (4.1), it holds that

$$\begin{aligned}\lim_{n\to\infty} \Bigg(&\Pr\left\{\frac{1}{n}\log\frac{1}{P_{V^n}(V^n)} \ge c_n\right\} \\ &+ \Pr\left\{\frac{1}{n}\log\frac{W^n(Y^n|X^n)}{P_{Y^n}(Y^n)} \le c_n + \gamma_n\right\}\Bigg) = 0, \end{aligned}\tag{4.9}$$

where Y^n is the channel output via W^n due to the channel input X^n.

Remark 4.2 (separation in general) @ The sufficient condition 2) in Theorem 4.3 means that the entropy spectrum of the source and the mutual information spectrum of the channel are asymptotically completely split with a vacant boundary of asymptotically vanishing width γ_n, and the former is placed to the left of the latter, where these two spectra may oscillate "synchronously" with n. In the case where such a separation condition 2) is satisfied, we can split reliable joint source-channel coding in two steps as follows (*separation* of source coding and channel coding): We first encode the source output V^n at the fixed-length coding rate $c_n = \frac{1}{n}\log M_n$ (M_n is the size of the message set $\mathcal{M}_n$), and then encode the output of the source encoder into the channel. The error probabilty ε_n for this two step coding is

upper bounded by the sum of the error probability of the fixed-length source coding (cf. Vembu, Verdú and Steinberg [9]; Han [11, Lemma 1.3.1]):

$$\Pr\left\{\frac{1}{n}\log\frac{1}{P_{V^n}(V^n)} \geq c_n\right\}$$

and the "maximum" error probability of the channel coding (cf. Feinstein [2], Ash [3], Han [11, Lemma 3.4.1]):

$$\Pr\left\{\frac{1}{n}\log\frac{W^n(Y^n|X^n)}{P_{Y^n}(Y^n)} \leq c_n+\gamma_n\right\} + e^{-n\gamma_n}.$$

It then follows from (4.9) that both of these two error probabilities vanish as n tends to ∞, where it should be noted that $e^{-n\gamma_n} \to 0$ as $n \to \infty$. Thus, we have $\lim_{n\to\infty} \varepsilon_n = 0$ to conclude that the source $\mathbf{V} = \{V^n\}_{n=1}^{\infty}$ is transmissible over the channel $\mathbf{W} = \{W^n\}_{n=1}^{\infty}$. This can be regarded as providing another proof of Theorem 4.1. □

Proof of Theorem 4.3:
2) ⇒ 1): For any joint probability distribution $P_{V^nX^n}$ for V^n and X^n, we have

$$\begin{aligned}
&\Pr\left\{\frac{1}{n}\log\frac{W^n(Y^n|X^n)}{P_{Y^n}(Y^n)} \leq \frac{1}{n}\log\frac{1}{P_{V^n}(V^n)} + \gamma_n\right\}\\
&\leq \quad \Pr\left\{\frac{1}{n}\log\frac{1}{P_{V^n}(V^n)} \geq c_n\right\}\\
&\qquad + \Pr\left\{\frac{1}{n}\log\frac{W^n(Y^n|X^n)}{P_{Y^n}(Y^n)} \leq c_n+\gamma_n\right\},
\end{aligned}$$

which together with (4.9) implies (4.8).

1) ⇒ 2)F Supposing that condition 1) holds, put

$$\alpha_n \equiv \Pr\left\{\frac{1}{n}\log\frac{W^n(Y^n|X^n)}{P_{Y^n}(Y^n)} \leq \frac{1}{n}\log\frac{1}{P_{V^n}(V^n)} + \gamma_n\right\}, \tag{4.10}$$

and moreover, with $\gamma_n' = \frac{\gamma_n}{4}$, $\delta_n = \max(\sqrt{\alpha_n}, e^{-n\gamma_n'})$, define

$$d_n = \sup\left\{R \,\middle|\, \Pr\left\{\frac{1}{n}\log\frac{1}{P_{V^n}(V^n)} \geq R\right\} > \delta_n\right\} - \gamma_n'. \tag{4.11}$$

Furthermore, define

$$S_n = \left\{ \mathbf{v} \in \mathcal{V}^n \,\middle|\, \frac{1}{n} \log \frac{1}{P_{V^n}(\mathbf{v})} \geq d_n \right\}, \tag{4.12}$$

$$\lambda_n^{(1)} = \Pr\left\{ V^n \in S_n \right\}, \quad \lambda_n^{(2)} = \Pr\left\{ V^n \notin S_n \right\}, \tag{4.13}$$

then the joint probability distribution $P_{V^n X^n Y^n}$ can be written as a mixture:

$$\begin{aligned} & P_{V^n X^n Y^n}(\mathbf{v}, \mathbf{x}, \mathbf{y}) \\ = \; & \lambda_n^{(1)} P_{\tilde{V}^n \tilde{X}^n \tilde{Y}^n}(\mathbf{v}, \mathbf{x}, \mathbf{y}) + \lambda_n^{(2)} P_{\overline{V}^n \overline{X}^n \overline{Y}^n}(\mathbf{v}, \mathbf{x}, \mathbf{y}), \end{aligned} \tag{4.14}$$

where $P_{\tilde{V}^n \tilde{X}^n \tilde{Y}^n}$, $P_{\overline{V}^n \overline{X}^n \overline{Y}^n}$ are the conditional probability distributions of $V^n X^n Y^n$ conditioned on $V^n \in S_n$, $V^n \notin S_n$, respectively. We notice here that the Markov chain property $V^n \to X^n \to Y^n$ implies $P_{\tilde{Y}^n | \tilde{X}^n} = P_{\overline{Y}^n | \overline{X}^n}$ $= W^n$ and the Markov chain properties

$$\tilde{V}^n \to \tilde{X}^n \to \tilde{Y}^n, \quad \overline{V}^n \to \overline{X}^n \to \overline{Y}^n.$$

We now rewrite (4.10) as

$$\begin{aligned} \alpha_n = \; & \lambda_n^{(1)} \Pr\left\{ \frac{1}{n} \log \frac{W^n(\tilde{Y}^n | \tilde{X}^n)}{P_{Y^n}(\tilde{Y}^n)} \leq \frac{1}{n} \log \frac{1}{P_{V^n}(\tilde{V}^n)} + \gamma_n \right\} \\ & + \lambda_n^{(2)} \Pr\left\{ \frac{1}{n} \log \frac{W^n(\overline{Y}^n | \overline{X}^n)}{P_{Y^n}(\overline{Y}^n)} \leq \frac{1}{n} \log \frac{1}{P_{V^n}(\overline{V}^n)} + \gamma_n \right\}. \end{aligned} \tag{4.15}$$

On the other hand, since (4.11), (4.12) lead to $\lambda_n^{(1)} > \delta_n \geq \sqrt{\alpha_n}$, it follows from (4.15) that

$$\Pr\left\{ \frac{1}{n} \log \frac{W^n(\tilde{Y}^n | \tilde{X}^n)}{P_{Y^n}(\tilde{Y}^n)} \leq \frac{1}{n} \log \frac{1}{P_{V^n}(\tilde{V}^n)} + \gamma_n \right\} \leq \sqrt{\alpha_n}. \tag{4.16}$$

Then, by the definition of $\tilde{V}^n$,

$$\frac{1}{n} \log \frac{1}{P_{V^n}(\tilde{V}^n)} \geq d_n,$$

and so from (4.16), we obtain

$$\Pr\left\{ \frac{1}{n} \log \frac{W^n(\tilde{Y}^n | \tilde{X}^n)}{P_{Y^n}(\tilde{Y}^n)} \leq d_n + \gamma_n \right\} \leq \sqrt{\alpha_n}. \tag{4.17}$$

Next, since it follows from (4.14) that

$$\begin{aligned} P_{Y^n}(\mathbf{y}) &= \lambda_n^{(1)} P_{\tilde{Y}^n}(\mathbf{y}) + \lambda_n^{(2)} P_{\overline{Y}^n}(\mathbf{y}) \\ &\geq \lambda_n^{(1)} P_{\tilde{Y}^n}(\mathbf{y}) \\ &\geq \delta_n P_{\tilde{Y}^n}(\mathbf{y}) \\ &\geq e^{-n\gamma_n'} P_{\tilde{Y}^n}(\mathbf{y}), \end{aligned}$$

we have

$$\frac{1}{n}\log\frac{1}{P_{Y^n}(\tilde{Y}^n)} \leq \frac{1}{n}\log\frac{1}{P_{\tilde{Y}^n}(\tilde{Y}^n)} + \gamma_n',$$

which is substituted into (4.17) to get

$$\Pr\left\{\frac{1}{n}\log\frac{W^n(\tilde{Y}^n|\tilde{X}^n)}{P_{\tilde{Y}^n}(\tilde{Y}^n)} \leq d_n + \gamma_n - \gamma_n'\right\} \leq \sqrt{\alpha_n}. \tag{4.18}$$

On the other hand, by the definition (4.11) of d_n,

$$\Pr\left\{\frac{1}{n}\log\frac{1}{P_{V^n}(V^n)} \geq d_n + 2\gamma_n'\right\} \leq \delta_n. \tag{4.19}$$

Set $c_n = d_n + 2\gamma_n'$ and note that $\alpha_n \to 0$, $\delta_n \to 0$ $(n \to \infty)$ and $\gamma_n' = \frac{\gamma_n}{4}$, then by (4.18), (4.19) we have

$$\lim_{n\to\infty}\left(\Pr\left\{\frac{1}{n}\log\frac{1}{P_{V^n}(V^n)} \geq c_n\right\} + \Pr\left\{\frac{1}{n}\log\frac{W^n(\tilde{Y}^n|\tilde{X}^n)}{P_{\tilde{Y}^n}(\tilde{Y}^n)} \leq c_n + \frac{1}{4}\gamma_n\right\}\right) = 0.$$

Finally, resetting $\tilde{X}^n\tilde{Y}^n$, $\frac{1}{4}\gamma_n$ as X^nY^n and γ_n, respectively, we conclude that condition 2), i.e., (4.9) holds. □

Having established an information-spectrum separation-like form of the sufficient condition (4.2) in Theorem 4.1, let us now turn to demonstrate several information-spectrum versions derived from the necessary condition (4.3) in Theorem 4.2.

Proposition 4.1 (Necessary conditions) The following two are necessary conditions for the transmissibility.

1) For *some* channel input $\mathbf{X} = \{X^n\}_{n=1}^{\infty}$ and for *any* sequence $\{\gamma_n\}_{n=1}^{\infty}$ satisfying condition (4.1), it holds that

$$\lim_{n\to\infty} \Pr\left\{\frac{1}{n}\log\frac{W^n(Y^n|X^n)}{P_{Y^n}(Y^n)} \le \frac{1}{n}\log\frac{1}{P_{V^n}(V^n)} - \gamma_n\right\} = 0, \tag{4.20}$$

where Y^n is the channel output via W^n due to the channel input X^n and $V^n \to X^n \to Y^n$.

2) For *any* sequence $\{\gamma_n\}_{n=1}^{\infty}$ satisfying condition (4.1) and for *some* channel input $\mathbf{X} = \{X^n\}_{n=1}^{\infty}$, it holds that

$$\lim_{n\to\infty} \Pr\left\{\frac{1}{n}\log\frac{W^n(Y^n|X^n)}{P_{Y^n}(Y^n)} \le \frac{1}{n}\log\frac{1}{P_{V^n}(V^n)} - \gamma_n\right\} = 0, \tag{4.21}$$

where Y^n is the channel output via W^n due to the channel input X^n and $V^n \to X^n \to Y^n$.

Proof: The necessity of condition 1) immediately follows from necessity condition (4.3) in Theorem 4.2. Moreover, it is also trivial to see that condition 1) implies condition 2) as an immediate logical consequence, and hence condition 2) is also a necessary condition. $\square$

The necessary condition 1) in Theorem 4.4 below is the same as condition 2) in Proposition 4.1. This is written here again in order to emphasize a pleasing duality between Theorem 4.3 and Theorem 4.4, which reflects on the duality between two fundamental Lemmas 3.1 and 3.2 .

Theorem 4.4 (Equivalence of necessary conditions) The following two conditions are equivalent:

1) For *any* sequence $\{\gamma_n\}_{n=1}^{\infty}$ satisfying condition (4.1) and for *some* channel input $\mathbf{X} = \{X^n\}_{n=1}^{\infty}$, it holds that

$$\lim_{n\to\infty} \Pr\left\{\frac{1}{n}\log\frac{W^n(Y^n|X^n)}{P_{Y^n}(Y^n)} \le \frac{1}{n}\log\frac{1}{P_{V^n}(V^n)} - \gamma_n\right\} = 0, \tag{4.22}$$

where Y^n is the channel output via W^n due to the channel input X^n and $V^n \to X^n \to Y^n$.

2) (**Domination**) For *any* sequence $\{\gamma_n\}_{n=1}^\infty$ satisfying condition (4.1) and for *some* channel input $\mathbf{X} = \{X^n\}_{n=1}^\infty$ and *some* sequence $\{c_n\}_{n=1}^\infty$, it holds that

$$\lim_{n\to\infty}\left(\Pr\left\{\frac{1}{n}\log\frac{1}{P_{V^n}(V^n)} \geq c_n\right\}\right.$$
$$\left. + \Pr\left\{\frac{1}{n}\log\frac{W^n(Y^n|X^n)}{P_{Y^n}(Y^n)} \leq c_n - \gamma_n\right\}\right) = 0, \qquad (4.23)$$

where Y^n is the channel output via W^n due to the channel input X^n.

Proof:

This theorem can be proved in the entirely same manner as in the proof of Theorem 4.3 with γ_n replaced by $-\gamma_n$. □

Remark 4.3 Originally, the definition of *domination* given by Vembu, Verdú and Steinberg [9] is not condition 2) in Theorem 4.4 but the following:

2′) (Domination) For *any* sequence $\{d_n\}_{n=1}^\infty$ and *any* sequence $\{\gamma_n\}_{n=1}^\infty$ satisfying condition (4.1), there exists *some* channel input $\mathbf{X} = \{X^n\}_{n=1}^\infty$ such that

$$\lim_{n\to\infty}\left(\Pr\left\{\frac{1}{n}\log\frac{1}{P_{V^n}(V^n)} \geq d_n\right\}\right.$$
$$\left. @ \times \Pr\left\{\frac{1}{n}\log\frac{W^n(Y^n|X^n)}{P_{Y^n}(Y^n)} \leq d_n - \gamma_n\right\}\right) = 0 \qquad (4.24)$$

holds, where Y^n is the channel output via W^n due to the channel input X^n. □

This necessary condition 2′) is implied by necessary condition 2) in Theorem 4.4. To see this, set

$$\alpha_n \equiv \Pr\left\{\frac{1}{n}\log\frac{1}{P_{V^n}(V^n)} \geq c_n\right\}, \qquad (4.25)$$

$$\beta_n \equiv \Pr\left\{\frac{1}{n}\log\frac{W^n(Y^n|X^n)}{P_{Y^n}(Y^n)} \leq c_n - \gamma_n\right\}, \qquad (4.26)$$

$$\kappa_n \equiv \Pr\left\{\frac{1}{n}\log\frac{1}{P_{V^n}(V^n)} \geq d_n\right\}, \qquad (4.27)$$

$$\mu_n \equiv \Pr\left\{\frac{1}{n}\log\frac{W^n(Y^n|X^n)}{P_{Y^n}(Y^n)} \leq d_n - \gamma_n\right\}. \qquad (4.28)$$

Then, we observe that $\kappa_n \le \alpha_n$ if $d_n \ge c_n$; and $\mu_n \le \beta_n$ if $d_n \le c_n$, and hence it follows from condition 2) that $\kappa_n \mu_n \le \alpha_n + \beta_n \to 0$ as n tends to ∞. Thus, condition 2) implies condition 2$'$), which means that condition 2) is strictly stronger than or equivalent to condition 2$'$) as necessary conditions for the transmissibility. It is not currently clear, however, whether both are equivalent or not. □

Remark 4.4 Condition 2) in Theorem 4.4 of this form is used later to directly prove Theorem 6.6 (separation theorem), while condition 2$'$) in Remark 4.3 of this form is irrelevant for this purpose. □

5 ε-Transmissibility Theorem

So far we have considered only the case where the error probability ε_n satisfies the condition $\lim_{n\to\infty} \varepsilon_n = 0$. However, we can relax this condition as follows:

$$\limsup_{n\to\infty} \varepsilon_n \le \varepsilon, \tag{5.1}$$

where ε is any constant such that $0 \le \varepsilon < 1$. (It is obvious that the special case with $\varepsilon = 0$ coincides with the case that we have considered so far.) We now say that the source $\mathbf{V}$ is *ε-transmissible* over the channel $\mathbf{W}$ when there exists an (n, ε_n) code satisfying condition (5.1).

Then, the same arguments as in the previous sections with due slight modifications lead to the following two theorems in parallel with Theorem 4.1 and Theorem 4.2, respectively:

Theorem 5.1 (ε-Direct theorem) Let $\mathbf{V} = \{V^n\}_{n=1}^\infty$, $\mathbf{W} = \{W^n\}_{n=1}^\infty$ be a general source and a general channel, respectively. If there exist *some* channel input $\mathbf{X} = \{X^n\}_{n=1}^\infty$ and *some* sequence $\{\gamma_n\}_{n=1}^\infty$ such that

$$\gamma_n > 0,\ \gamma_n \to 0 \text{ and } n\gamma_n \to \infty \quad (n \to \infty) \tag{5.2}$$

for which it holds that

$$\limsup_{n\to\infty} \Pr\left\{ \frac{1}{n} \log \frac{W^n(Y^n|X^n)}{P_{Y^n}(Y^n)} \le \frac{1}{n} \log \frac{1}{P_{V^n}(V^n)} + \gamma_n \right\} \le \varepsilon, \tag{5.3}$$

then the source $\mathbf{V} = \{V^n\}_{n=1}^\infty$ is ε-transmissible over the channel $\mathbf{W} = \{W^n\}_{n=1}^\infty$, where Y^n is the channel output via W^n due to the channel input

X^n and $V^n \to X^n \to Y^n$. $\square$

Theorem 5.2 (ε-Converse theorem) Suppose that a general source $\mathbf{V} = \{V^n\}_{n=1}^{\infty}$ is ε-transmissible over a general channel $\mathbf{W} = \{W^n\}_{n=1}^{\infty}$, and let the channel input be $\mathbf{X} = \{X^n \equiv \varphi_n(V^n)\}_{n=1}^{\infty}$ where $\varphi_n : \mathcal{V}^n \to \mathcal{X}^n$ is the channel encoder. Then, for *any* sequence $\{\gamma_n\}_{n=1}^{\infty}$ satisfying condition (5.2), it holds that

$$\limsup_{n\to\infty} \Pr\left\{\frac{1}{n}\log\frac{W^n(Y^n|X^n)}{P_{Y^n}(Y^n)} \le \frac{1}{n}\log\frac{1}{P_{V^n}(V^n)} - \gamma_n\right\} \le \varepsilon, \tag{5.4}$$

where Y^n is the channel output via W^n due to the channel input X^n and $V^n \to X^n \to Y^n$. $\square$

Remark 5.1 It should be noted here that such a sufficient condition (5.3) as well as such a necessary condition (5.4) for the ε-transmissibility cannot actually be derived in the way of generalizing the strict domination in (4.9) and the domination in (4.23). It should be noted also that, under the ε-transmissibility criterion, joint source-channel coding is beyond the separation principle. $\square$

6 Separation Theorems of the Traditional Type

Thus far we have investigated the joint source-channel coding problem from the viewpoint of information spectra and established the fundamental theorems (Theorems 4.1$\sim$4.4). These results are of seemingly different forms from separation theorems of the traditional type. Then, it would be natural to ask a question how the separation principle of the information spectrum type is related to separation theorems of the traditional type. In this section we address this question.

To do so, we first need some preparation. We denote by $R_f(\mathbf{V})$ the infimum of all achievable fixed-length coding rates for a general source $\mathbf{V} = \{V^n\}_{n=1}^{\infty}$ (as for the formal definition, see Han and Verdú [7], Han [11, Definitions 1.1.1, 1.1.2]), and denote by $C(\mathbf{W})$ the capacity of a general channel $\mathbf{W} = \{W^n : \mathcal{X}^n \to \mathcal{Y}^n\}_{n=1}^{\infty}$ (as for the formal definition, see Han and Verdú [7], Han [11, Definitions 3.1.1, 3.1.2]). First, $R_f(\mathbf{V})$ is characterized as

Theorem 6.1 (Han and Verdú [7], Han[11])

$$R_f(\mathbf{V}) = \overline{H}(\mathbf{V}), \tag{6.1}$$

where [§]

$$\overline{H}(\mathbf{V}) = \text{p-}\limsup_{n\to\infty} \frac{1}{n} \log \frac{1}{P_{V^n}(V^n)}. \tag{6.2}$$

Next, let us consider about the characterization of $C(\mathbf{W})$. Given a general channel $\mathbf{W} = \{W^n\}_{n=1}^\infty$ and its input $\mathbf{X} = \{X^n\}_{n=1}^\infty$, let $\mathbf{Y} = \{Y^n\}_{n=1}^\infty$ be the output due to the input $\mathbf{X} = \{X^n\}_{n=1}^\infty$ via the channel $\mathbf{W} = \{W^n\}_{n=1}^\infty$. Define

Definition 6.1

$$\underline{I}(\mathbf{X};\mathbf{Y}) = \text{p-}\liminf_{n\to\infty} \frac{1}{n} \log \frac{W^n(Y^n|X^n)}{P_{Y^n}(Y^n)}. \tag{6.3}$$

Then, the capacity $C(\mathbf{W})$ is characterized as follows.

Theorem 6.2 (Verdú and Han [8], Han[11])

$$C(\mathbf{W}) = \sup_{\mathbf{X}} \underline{I}(\mathbf{X};\mathbf{Y}), \tag{6.4}$$

where $\sup_{\mathbf{X}}$ means the supremum over all possible inputs $\mathbf{X}$. □

With these preparations, let us turn to the separation theorem problem of the traditional type. A general source $\mathbf{V} = \{V^n\}_{n=1}^\infty$ is said to be *information-stable* (cf. Dobrushin [4], Pinsker [5]) if

$$\frac{\frac{1}{n}\log\frac{1}{P_{V^n}(V^n)}}{H_n(V^n)} \to 1 \quad \text{in prob.}, \tag{6.5}$$

where $H_n(V^n) = \frac{1}{n}H(V^n)$ and $H(V^n)$ stands for the entropy of V^n (cf. Cover and Thomas [13]). Moreover, a general channel $\mathbf{W} = \{W^n\}_{n=1}^\infty$ is

[§]For an arbitrary sequence of real-valued random variables $\{Z_n\}_{n=1}^\infty$, we define the following notions (cf. Han and Verdú [7], Han[11]): $\text{p-}\limsup_{n\to\infty} Z_n \equiv \inf\{\alpha \mid \lim_{n\to\infty} \Pr\{Z_n > \alpha\} = 0\}$ (the *limit superior in probability*), and $\text{p-}\liminf_{n\to\infty} Z_n \equiv \sup\{\beta \mid \lim_{n\to\infty} \Pr\{Z_n < \beta\} = 0\}$ (the *limit inferior in probability*).

said to be *information-stable* (cf. Dobrushin [4], Pinsker [5], Hu [6]) if there exists a channel input $\mathbf{X} = \{X^n\}_{n=1}^{\infty}$ such that

$$\frac{\frac{1}{n}\log\frac{W(Y^n|X^n)}{P_{Y^n}(Y^n)}}{C_n(W^n)} \to 1 \quad \text{in prob.,} \tag{6.6}$$

where

$$C_n(W^n) = \sup_{X^n} \frac{1}{n} I(X^n; Y^n),$$

and Y^n is the channel output via W^n due to the channel input X^n; and $I(X^n;Y^n)$ is the mutual information between X^n and Y^n (cf. Cover and Thomas [13]). Then, we can summarize a typical separation theorem of the traditional type as follows.

Theorem 6.3 (Dobrushin [4], Pinsker [5]) Let the channel $\mathbf{W} = \{W^n\}_{n=1}^{\infty}$ be information-stable and suppose that the limit $\lim_{n\to\infty} C_n(W^n)$ exists, or, let the source $\mathbf{V} = \{V^n\}_{n=1}^{\infty}$ be information-stable and suppose that the limit $\lim_{n\to\infty} H_n(V^n)$ exists. Then, the following two statements hold:

1) If $R_f(\mathbf{V}) < C(\mathbf{W})$, then the source $\mathbf{V}$ is transmissible over the channel $\mathbf{W}$. In this case, we can separate the source coding and the channel coding.

2) If the source $\mathbf{V}$ is transmissible over the channel $\mathbf{W}$, then it must hold that $R_f(\mathbf{V}) \le C(\mathbf{W})$. □

In order to generalize Theorem 6.3, we need to introduce the concept of *optimistic* coding. The "optimistic" standpoint means that we evaluate the coding reliability with error probability $\liminf_{n\to\infty} \varepsilon_n = 0$ (that is, $\varepsilon_n < \forall\varepsilon$ *for infinitely many* n). In contrast with this, the standpoint that we have taken so far is called *pessimistic* with error probability $\lim_{n\to\infty} \varepsilon_n = 0$ (that is, $\varepsilon_n < \forall\varepsilon$ *for all sufficiently large* n).

The following one concerns the optimistic source coding with any general source $\mathbf{V}$.

Definition 6.2 (Optimistic achievability for source coding)

Rate R is optimistically achievable $\overset{\text{def}}{\Longleftrightarrow}$ There exists an (n, M_n, ε_n)- source code satisfying $\liminf\limits_{n\to\infty} \varepsilon_n = 0$ and

$$\limsup_{n\to\infty} \frac{1}{n}\log M_n \le R,$$

where $\frac{1}{n}\log M_n$ is the coding rate per source letter (see, e.g., Han [11, Section 1.1]).

Definition 6.3 (Optimistic achievable fixed-length coding rate)

$$\underline{R}_f(\mathbf{V}) = \inf\left\{R \mid R \text{ is optimistically achievable}\right\}.$$

Then, for any general source $\mathbf{V} = \{V^n\}_{n=1}^{\infty}$ we have:

Theorem 6.4 (Chen and Alajaji [14])

$$\underline{R}_f(\mathbf{V}) = \inf\left\{R \,\middle|\, \liminf_{n\to\infty} \Pr\left\{\frac{1}{n}\log\frac{1}{P_{V^n}(V^n)} \geq R\right\} = 0\right\}. \tag{6.7}$$

On the other hand, the next one concerns the optimistic channel capacity.

Definition 6.4 (Optimistic achievability for channel coding)

Rate R is optimistically achievable $\overset{\text{def}}{\Longleftrightarrow}$ There exists an (n, M_n, ε_n)-channel code satisfying $\liminf_{n\to\infty}\varepsilon_n = 0$ and $\liminf_{n\to\infty}\frac{1}{n}\log M_n \geq R$,

where $\frac{1}{n}\log M_n$ is the coding rate per channel use (see, e.g., Han [11, Section 3.1]).

Definition 6.5 (Optimistic channel capacity)

$$\overline{C}(\mathbf{W}) = \sup\left\{R \mid R \text{ is optimistically achievable}\right\}.$$

Then, with a general channel $\mathbf{W} = \{W^n\}_{n=1}^{\infty}$ we have

Theorem 6.5 (Chen and Alajaji [14])

$$\begin{aligned}&\overline{C}(\mathbf{W})\\ &= \sup_{\mathbf{X}}\sup\left\{R \,\middle|\, \liminf_{n\to\infty} \Pr\left\{\frac{1}{n}\log\frac{W^n(Y^n|X^n)}{P_{Y^n}(Y^n)} \leq R\right\} = 0\right\},\end{aligned} \tag{6.8}$$

where Y^n is the output due to the input $\mathbf{X} = \{X^n\}_{n=1}^{\infty}$. □

Remark 6.1 It is not difficult to check that, in parallel with Theorem 6.4 and Theorem 6.5, Theorem 6.1 and Theorem 6.2 can be rewritten as

$$R_f(\mathbf{V}) = \inf\left\{R \,\middle|\, \lim_{n\to\infty} \Pr\left\{\frac{1}{n}\log\frac{1}{P_{V^n}(V^n)} \geq R\right\} = 0\right\}, \tag{6.9}$$

$$C(\mathbf{W}) = \sup_{\mathbf{X}} \sup\left\{R \,\middle|\, \lim_{n\to\infty} \Pr\left\{\frac{1}{n}\log\frac{W^n(Y^n|X^n)}{P_{Y^n}(Y^n)} \leq R\right\} = 0\right\}, \tag{6.10}$$

from which, together with Theorem 6.4 and Theorem 6.5, it immediately follows that

$$C(\mathbf{W}) \leq \overline{C}(\mathbf{W}), \tag{6.11}$$

$$\underline{R}_f(\mathbf{V}) \leq R_f(\mathbf{V}). \tag{6.12}$$

Now, we have:

Theorem 6.6 Let $\mathbf{W} = \{W^n\}_{n=1}^{\infty}$ be a general channel and $\mathbf{V} = \{V^n\}_{n=1}^{\infty}$ be a general source. Then, the following two statements hold:

1) If $R_f(\mathbf{V}) < C(\mathbf{W})$, then the source $\mathbf{V}$ is transmissible over the channel $\mathbf{W}$. In this case, we can separate the source coding and the channel coding.

2) If the source $\mathbf{V}$ is transmissible over the channel $\mathbf{W}$, then it must hold that

$$\underline{R}_f(\mathbf{V}) \leq C(\mathbf{W}), \tag{6.13}$$

$$R_f(\mathbf{V}) \leq \overline{C}(\mathbf{W}). \tag{6.14}$$

Remark 6.2 As was mentioned in Remark 4.4, we use Theorem 4.4 in order to prove (6.13) and (6.14), where inequality (6.14) was shown in a rather roundabout manner by Vembu, Verdú and Steinberg [9] (invoking Domination 2′) in Remark 4.3 instead of Domination 2) in Theorem 4.4). □

Proof of Theorem 6.6.
1): Since $R_f(\mathbf{V}) = \overline{H}(\mathbf{V})$, $C(\mathbf{W}) = \sup_{\mathbf{X}} \underline{I}(\mathbf{X};\mathbf{Y})$ by Theorem 6.1 and Theorem 6.2, the inequality $R_f(\mathbf{V}) < C(\mathbf{W})$ implies that condition 2) in Theorem 4.3 holds for $\mathbf{X} = \{X^n\}_{n=1}^{\infty}$ attaining the supremum $\sup_{\mathbf{X}} \underline{I}(\mathbf{X};\mathbf{Y})$

with, for example, $c_n = \frac{1}{2}(R_f(\mathbf{V}) + C(\mathbf{W}))$. Therefore, the source $\mathbf{V}$ is transmissible over the channel $\mathbf{W}$.

2): If the source $\mathbf{V}$ is transmissible over the channel $\mathbf{W}$, then condition 2) in Theorem 4.4 holds with some $\{c_n\}_{n=1}^{\infty}$, i.e.,

$$\lim_{n\to\infty} \Pr\left\{\frac{1}{n}\log\frac{1}{P_{V^n}(V^n)} \geq c_n\right\} = 0, \tag{6.15}$$

$$\lim_{n\to\infty} \Pr\left\{\frac{1}{n}\log\frac{W^n(Y^n|X^n)}{P_{Y^n}(Y^n)} \leq c_n - \gamma_n\right\} = 0. \tag{6.16}$$

Since $\lim_{n\to\infty}\gamma_n = 0$, these two conditions with any small constant $\delta > 0$ lead us to the following formulas:

$$\liminf_{n\to\infty} \Pr\left\{\frac{1}{n}\log\frac{1}{P_{V^n}(V^n)} \geq \liminf_{n\to\infty} c_n + \delta\right\} = 0, \tag{6.17}$$

$$\lim_{n\to\infty} \Pr\left\{\frac{1}{n}\log\frac{1}{P_{V^n}(V^n)} \geq \limsup_{n\to\infty} c_n + \delta\right\} = 0, \tag{6.18}$$

$$\lim_{n\to\infty} \Pr\left\{\frac{1}{n}\log\frac{W^n(Y^n|X^n)}{P_{Y^n}(Y^n)} \leq \liminf_{n\to\infty} c_n - \delta\right\} = 0, \tag{6.19}$$

$$\liminf_{n\to\infty} \Pr\left\{\frac{1}{n}\log\frac{W^n(Y^n|X^n)}{P_{Y^n}(Y^n)} \leq \limsup_{n\to\infty} c_n - \delta\right\} = 0. \tag{6.20}$$

Then, Theorem 6.4 and (6.17) imply that $\underline{R}_f(\mathbf{V}) \leq \liminf_{n\to\infty} c_n$, whereas (6.19) implies that $\underline{I}(\mathbf{X};\mathbf{Y}) \geq \liminf_{n\to\infty} c_n$. Therefore, by Theorem 6.2 we have

$$\underline{R}_f(\mathbf{V}) \leq \liminf_{n\to\infty} c_n \leq \underline{I}(\mathbf{X};\mathbf{Y}) \leq \sup_{\mathbf{X}} \underline{I}(\mathbf{X};\mathbf{Y}) = C(\mathbf{W}).$$

On the other hand, (6.18) implies that $\overline{H}(\mathbf{V}) \leq \limsup_{n\to\infty} c_n$. Furthermore, (6.20) together with Theorem 6.5 gives us

$$\overline{H}(\mathbf{V}) \leq \limsup_{n\to\infty} c_n \leq \overline{C}(\mathbf{W}).$$

Finally, note that $R_f(\mathbf{V}) = \overline{H}(\mathbf{V})$ by Theorem 6.1. □

We are now interested in the problem of what conditions are needed to attain equalities $\underline{R}_f(\mathbf{V}) = R_f(\mathbf{V})$ and/or $\overline{C}(\mathbf{W}) = C(\mathbf{W})$ in Theorem 6.6 and so on. To see this, we need the following four definitions:

Definition 6.6 A general source $\mathbf{V} = \{V^n\}_{n=1}^{\infty}$ is said to satisfy the *strong converse property* if

$$\overline{H}(\mathbf{V}) = \underline{H}(\mathbf{V})$$

holds (as for the operational meaning, refer to Han [11]), where

$$\underline{H}(\mathbf{V}) = \text{p-}\liminf_{n\to\infty} \frac{1}{n} \log \frac{1}{P_{V^n}(V^n)}.$$

Definition 6.7 A general channel $\mathbf{W} = \{W^n\}_{n=1}^{\infty}$ is said to satisfy the *strong converse property* if

$$\sup_{\mathbf{X}} \underline{I}(\mathbf{X};\mathbf{Y}) = \sup_{\mathbf{X}} \overline{I}(\mathbf{X};\mathbf{Y}) \tag{6.21}$$

holds (as for the operational meaning, refer to Han [11], Verdú and Han [8]), where

$$\overline{I}(\mathbf{X};\mathbf{Y}) = \text{p-}\limsup_{n\to\infty} \frac{1}{n} \log \frac{W^n(Y^n|X^n)}{P_{Y^n}(Y^n)}.$$

Definition 6.8 A general source $\mathbf{V} = \{V^n\}_{n=1}^{\infty}$ is said to satisfy the *semi-strong converse property* if for *all* divergent subsequences $\{n_i\}_{n=1}^{\infty}$ of positive integers such that $n_1 < n_2 < \cdots \to \infty$ it holds that

$$\text{p-}\limsup_{i\to\infty} \frac{1}{n_i} \log \frac{1}{P_{V^{n_i}}(V^{n_i})} = \overline{H}(\mathbf{V}). \tag{6.22}$$

Definition 6.9 A general channel $\mathbf{W} = \{W^n\}_{n=1}^{\infty}$ is said to satisfy the *semi-strong converse property* if for *all* divergent subsequences $\{n_i\}_{n=1}^{\infty}$ of positive integers such that $n_1 < n_2 < \cdots \to \infty$ it holds that

$$\text{p-}\liminf_{i\to\infty} \frac{1}{n_i} \log \frac{W^{n_i}(Y^{n_i}|X^{n_i})}{P_{Y^{n_i}}(Y^{n_i})} \le \sup_{\mathbf{X}} \underline{I}(\mathbf{X};\mathbf{Y}), \tag{6.23}$$

where Y^n is the channel output via W^n due to the channel input X^n. □

With these definitions, we have the following lemmas:

Lemma 6.1

1) The information-stability of a source $\mathbf{V}$ (resp. a channel $\mathbf{W}$) with the limit implies the strong converse property of $\mathbf{V}$ (resp. $\mathbf{W}$).

2) The strong converse property of a source $\mathbf{V}$ (resp. a channel $\mathbf{W}$) implies the semi-strong converse property of $\mathbf{V}$ (resp. $\mathbf{W}$). □

Lemma 6.2

1) A general source $\mathbf{V}$ satisfies the semi-strong converse property if and only if

$$\underline{R}_f(\mathbf{V}) = R_f(\mathbf{V}). \tag{6.24}$$

2) A general channel $\mathbf{W}$ satisfies the semi-strong converse property if and only if

$$\overline{C}(\mathbf{W}) = C(\mathbf{W}). \tag{6.25}$$

Proof: It is obvious in view of Theorem 6.4, Theorem 6.5 and Remark 6.1. □

Remark 6.3 An operational equivalent of the notion of semi-strong converse property is found in Vembu, Verdú and Steinberg [9]. Originally, Csiszár and Körner [12] posed two operational standpoints in source coding and channel coding, i.e., the *pessimistic standpoint* and the *optimistic standpoint*. In their terminology, Lemma 6.2 states that, for source coding, the semi-strong convserse property is equivalent to the statement that both the pessimistic standpoint and the optiimistic standpoint result in the same infimum of all achievable fixed-length source coding rates; similarly, for channel coding, the semi-strong convserse property is equivalent to the claim that both the pessimistic standpoint and the optimistic standpoint result in the same supremum of all achievable channel coding rates. □

Thus, Theorem 6.6 together with Lemma 6.2 immediately yields the following stronger separation theorem of the traditional type:

Theorem 6.7 Let either a general source $\mathbf{V} = \{V^n\}_{n=1}^{\infty}$ or a general channel $\mathbf{W} = \{W^n\}_{n=1}^{\infty}$ satisfy the semi-strong converse property. Then, the following two statements hold:

1) If $R_f(\mathbf{V}) < C(\mathbf{W})$, then the source $\mathbf{V}$ is transmissible over the channel $\mathbf{W}$. In this case, we can separate the source coding and the channel coding.

2) If the source $\mathbf{V}$ is transmissible over the channel $\mathbf{W}$, then it must hold that $R_f(\mathbf{V}) \le C(\mathbf{W})$. $\square$

Example 6.1 Theorem 6.3 is an immediate consequence of Theorem 6.7 together with Lemma 6.1. $\square$

Example 6.2 Let us consider two different stationary memoryless sources $\mathbf{V}_1 = \{V_1^n\}_{n=1}^\infty$, $\mathbf{V}_2 = \{V_2^n\}_{n=1}^\infty$ with countably infinite source alphabet $\mathcal{V}$, and define its *mixed* source $\mathbf{V} = \{V^n\}_{n=1}^\infty$ by

$$P_{V^n}(\mathbf{v}) = \alpha_1 P_{V_1^n}(\mathbf{v}) + \alpha_2 P_{V_2^n}(\mathbf{v}) \quad (\mathbf{v} \in \mathcal{V}^n),$$

where α_1, α_2 are positive constants such that $\alpha_1 + \alpha_2 = 1$. Then, this mixed source $\mathbf{V} = \{V^n\}_{n=1}^\infty$ satisfies the semi-strong converse property but neither the strong converse property nor the information-stability.

Similarly, let us consider two different stationary memoryless channels $\mathbf{W}_1 = \{W_1^n\}_{n=1}^\infty$, $\mathbf{W}_2 = \{W_2^n\}_{n=1}^\infty$ with arbitrary abstract input and output alphabets $\mathcal{X}, \mathcal{Y}$, and define its *mixed* channel $\mathbf{W} = \{W^n\}_{n=1}^\infty$ by

$$W^n(\mathbf{y}|\mathbf{x}) = \alpha_1 W_1^n(\mathbf{y}|\mathbf{x}) + \alpha_2 W_2^n(\mathbf{y}|\mathbf{x}) \quad (\mathbf{x} \in \mathcal{X}^n, \mathbf{y} \in \mathcal{Y}^n).$$

Then, this mixed channel $\mathbf{W} = \{W^n\}_{n=1}^\infty$ satisfies the semi-strong converse property but neither the strong converse property nor the information-stability.

Thus, in these mixed cases the separation theorem holds. $\square$

References

[1] C. E. Shannon, "A mathematical theory of communication," *Bell System Technical Journal*, vol.27, pp.379-423, pp. 623-656, 1948

[2] A. Feinstein, "A new basic theorem of information theory," *IRE Trans. PGIT*, vol.4, pp.2-22, 1954

[3] R.B. Ash, *Information Theory*, Interscience Publishers, New York, 1965

[4] R. L. Dobrushin, "A general formulation of the fundamental Shannon theorem in information theory," *Uspehi Mat. Acad. Nauk. SSSR*, vol.40, pp.3-104, 1959: Translation in *Transactions of American Mathematical Society*, Series 2, vol.33, pp.323-438, 1963

[5] M. S. Pinsker, *Information and Information Stability of Random Variables and Processes*, Holden-Day, San Francisco, 1964

[6] G. D. Hu, "On Shannon theorem and its converse for sequence of communication schemes in the case of abstract random variables," in *Trans. 3rd Prague Conference on Information Theory, Statistical Decision Functions, Random Processes*, Czechslovak Academy of Sciences, Prague, pp. 285-333, 1964

[7] T.S. Han and S. Verdú, "Approximation theory of output statistics," *IEEE Transactions on Information Theory*, vol.IT-39, no.3, pp. 752-772, 1993

[8] S. Verdú and T.S. Han, "A general formula for channel capacity," *IEEE Transactions on Information Theory*, vol.IT-40, no.4, pp.1147-1157, 1994

[9] S. Vembu, S. Verdú and Y. Steinberg, "The source-channel separation theorem revisited," *IEEE Transactions on Information Theory*, vol.IT-41, no.1, pp. 44-54, 1995

[10] S. Verdú and T. S. Han, "The role of the asymptotic equipartition property in noiseless source coding," *IEEE Transactions on Information Theory*, vol.IT-43, no.3, pp.847-857, 1997

[11] T. S. Han, *Information-Spectrum Methods in Information Theory*, Springer Verlag, New York, 2003

[12] I. Csiszár and J. Körner, *Information Theory: Coding Theorems for Discrete Memoryless Systems*, Academic Press, New York, 1981

[13] T. M. Cover and J. Thomas, *Elements of Information Theory*, Wiley, New York, 1991

[14] P.N. Chen and F. Alajaji, "Optimistic Shannon coding theorems for arbitrary single-user systems," *IEEE Transactions on Information Theory*, IT-45, pp. 2623-2629, 1999